# Estimation of Hydrologic Parameters under Unsaturated Flow Conditions using the Markov Chain Monte Carlo Simulations

Zhiming Lu*
Bruce A. Robinson

Hydrology, Geochemistry, and Geology Group (EES-6)
MS T003, Los Alamos National Laboratory
Los Alamos, NM 87545

David Higdon

Statistical Sciences Group (D-1)
F600, Los Alamos National Laboratory
Los Alamos, NM 87545

*Corresponding Author: (505) 665-2126. Fax (505) 665-8737. zhiming@lanl.gov

For submission to Vadose Zone Journal

January 31, 2006

## Abstract

An inversion technique based on the Markov Chain Monte Carlo (MCMC) method is developed for estimating statistics of the infiltration rate and characterizing unsaturated porous media, using measurements on hydraulic properties, pressure head, water content, and solute concentration or travel time. The MCMC realizations are taken from a posterior distribution, which incorporates all available data. The method is first tested using a synthetic dataset, which demonstrates that it can reproduce the true dataset very well. The method is then applied to the interpretation of water content data from boreholes around the Los Alamos National Laboratory (LANL) for the purpose of obtaining estimates of the local infiltration rate. Infiltration rates obtained from modeling of water content data from wells in Los Alamos and Mortandad Canyons are in general agreement with previous estimates. However, this method also provides reasonable estimates of uncertainty for wells in a variety of topographic settings, partially because it takes into account heterogeneity in the medium properties. Numerical experiments with the method illustrate that including small-scale heterogeneity is important for improved matches to the data.

## 1. Introduction

Accurately predicting flow and solute transport in subsurface requires accurate and detailed estimation of hydrologic parameters as well as uncertainties associated with the estimation. Parameter identification has been a major research area in the last two decades, and many inverse methods have been developed (Yeh, 1986; Ginn and Cushman, 1990; McLaughlin and Townley, 1996; Carrera et al., 2005). Zimmerman et al. (1998) gave a comprehensive comparison on seven different geostatistically based methods. While many different methods have been used in characterizing saturated porous media, applications of these methods to inverse problems under unsaturated flow conditions is in very limited. Yeh and Zhang (1996) and Li and Yeh (1999) developed geostatistically based sequential cokriging methods, which incorporate measurements on the direct measurements of hydraulic conductivity and pore-size distribution parameter as well as measurements on the dependent variables such as pressure head, solute concentration or travel time. Vrugt et al. (2004) proposed an inverse model for large-scale spatially distributed vadose zone properties using global optimization.

The Markov Chain Monte Carlo method (MCMC) is a powerful technique in sampling parameter space. The method has been used in a number of applications such as history data match (Oliver et al, 1997), charactering saturated porous media (Lu et al., 2004;), data integration (Lee at el., 2000; Efendiev et al., 2005), and geochemical characterization (Chen et al., 2004). One of the advantages of the MCMC method is that it will generate samples from the correct posterior probability density function (PDF). In this study, we developed a Markov Chain Monte Carlo method (MCMC) for characterizing hydrologic properties under unsaturated flow conditions and applied this method to estimation of hydrologic parameters (infiltration

72 rates, saturated hydraulic conductivity, pore-size distribution parameter, and the fitting parameter
73 of the von Genuchton constitutive relationship) at the vicinity of the Los Alamos National
74 Laboratory (LANL). The results show that the method provides reasonable estimates of
75 uncertainty of the infiltration rate for wells in a variety of topographic settings.

76 **2. Statement of the Problem**

77 We consider transient flow in variably saturated porous media satisfying the following continuity
78 equation and Darcy's law:

79 $$-\nabla \cdot \mathbf{q}(\mathbf{x},t) + g(\mathbf{x},t) = C_s(\psi)\frac{\partial \psi(\mathbf{x},t)}{\partial t}, \qquad (1)$$

80 $$\mathbf{q}(\mathbf{x},t) = -K[\psi]\nabla[\psi(\mathbf{x},t) + x_1], \qquad (2)$$

81 subject to initial and boundary conditions

82 $$\psi(\mathbf{x},0) = \Psi_0(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \qquad (3)$$

83 $$\psi(\mathbf{x},t) = \Psi(\mathbf{x},t), \qquad \mathbf{x} \in \Gamma_D \qquad (4)$$

84 $$\mathbf{q}(\mathbf{x},t) \cdot \mathbf{n}(\mathbf{x}) = Q(\mathbf{x},t), \qquad \mathbf{x} \in \Gamma_N, \qquad (5)$$

85 where $\boldsymbol{q}$ is the specific discharge (flux), $\psi(\boldsymbol{x},t) + x_1$ is the total head, $\psi$ is the pressure head,
86 $\Psi_0(\boldsymbol{x})$ is the initial pressure head in the domain $\Omega$, $\Psi(\boldsymbol{x},t)$ is the prescribed head on Dirichlet
87 boundary segments $\Gamma_D$, $Q(\boldsymbol{x},t)$ is the prescribed flux across Neumann boundary segments $\Gamma_N$,
88 $\boldsymbol{n}(\boldsymbol{x})=(n_1,\cdots, n_d)^T$ is an outward unit vector normal to the boundary, $C[\psi]= d\theta/\mathrm{d}\psi$ is the specific
89 moisture capacity, $\theta$ is the volumetric water content, and $K[\psi]$ is the unsaturated hydraulic

90 conductivity (assumed to be isotropic locally). Both $C$ and $K$ are functions of pressure head and
91 soil properties at $\boldsymbol{x}$. For convenience, they will be written as $C(\boldsymbol{x},t)$ and $K(\boldsymbol{x},t)$ in the sequel. The
92 elevation $x_1$ is directed vertically upward. In these coordinates, recharge has a negative sign.

93 It is clear that models are needed to describe the constitutive relationships of K versus $\psi$ and $\theta_e$
94 versus $\psi$ when the flow is unsaturated. No universal models are available for the constitutive
95 relationships. Instead, several empirical models are usually used, including the Gardner-Russo
96 model [Gardner, 1958; Russo, 1988], the Brooks-Corey model [Brooks and Corey, 1964], and
97 the van Genuchten-Mualem model [van Genuchten, 1980]. Most analytical solutions of the
98 deterministic unsaturated flow equations and most stochastic analyses used the Gardner-Russo
99 model because of its simplicity. However, it is generally accepted that the more complex van
100 Genuchten-Mualem and Brooks-Corey models may perform better than the simple Gardner-
101 Russo model in describing measured data of $K(\psi)$ and $\theta_e(\psi)$. In this study, we use the van
102 Genuchten-Mualem model:

103 $$K(\mathbf{x},t) = K_s(\mathbf{x})\sqrt{S(\mathbf{x},t)}\{1-[1-S^{1/m}(\mathbf{x},t)]^m\}^2 \tag{7}$$

104 $$S(\mathbf{x},t) = \{1+[-\alpha(\mathbf{x})\psi(\mathbf{x},t)]^n\}^{-m}, \tag{8}$$

105 where $\psi \leq 0$. In the above, $S(\mathbf{x},t)= \theta_e/(\theta_s-\theta_r)$ is the effective saturation, $\theta_r$ is the residual
106 (irreducible) water content, $\theta_s$ is the saturated water content, $\alpha$ is the pore-size distribution
107 parameter, $n$ is a fitting parameter, and $m = 1 - 1/n$. With (8), $C_s(\mathbf{x},t) = d\theta_e/\mathrm{d}\psi$ can be expressed
108 explicitly as

109 $$C_s(\mathbf{x},t) = \alpha(\mathbf{x})[n(\mathbf{x})-1](\theta_s-\theta_r)S^{1/m}(\mathbf{x},t)[1-S^{1/m}(\mathbf{x},t)]^m \tag{9}$$

In this study, $\theta_s$ and $\theta_r$ are assumed to be deterministic as their variabilities are likely to be small compared to that of the effective water content $\theta_e$ [*Russo and Bouton,* 1992], while the saturated hydraulic conductivity $K_s$, the pore size distribution parameter $\alpha$, and the fitting parameter $n$ are treated as random functions [*Lu and Zhang*, 2002]. We assume that the log-transformed saturated hydraulic conductivity $f(\boldsymbol{x}) = \ln K_s(\boldsymbol{x})$, the log-transformed pore size distribution parameter $\beta(\boldsymbol{x}) = \ln \alpha(\boldsymbol{x})$, and $\mu(\boldsymbol{x}) = \ln [n(\boldsymbol{x})-1]$ follow normal distributions. Now suppose that there are $m_f$ direct measurements on the log hydraulic conductivity, $f_i$, $i = \overline{1, m_f}$, $m_\beta$ measurements on the pore-size distribution parameter, $\beta_i$, $i = \overline{1, m_\beta}$, and $m_\mu$ measurements on the fitting parameter, $\mu_i$, $i = \overline{1, m_\mu}$. The sampling locations for three different kinds of direct measurements may be different.

The transport of a nonreactive solute in the heterogeneous porous media can be described by the classical convection-dispersion transport equation [Bear, 1972]:

$$\nabla\left[D_{ij}\cdot\nabla C(\mathbf{x},t)\right]-v\cdot\nabla C(\mathbf{x},t)=\theta\frac{\partial C(\mathbf{x},t)}{\partial t} \tag{10}$$

subject to appropriate initial and boundary conditions. Here $C$ is the nonreactive solute concentration, $D_{ij}$ is the dispersion coefficient tensor, and $\boldsymbol{v}$ is the seepage velocity, which can be computed from specific discharge $\boldsymbol{q}$, as solved from (2).

An alternative way to characterize nonreactive transport is to record the position of a particle at time $t$ that originates from position $\mathbf{a}$ at time $t = t_0$ and is described by the following kinetic equation

129 $$\frac{d\mathbf{X}(t;\mathbf{a})}{dt} = \mathbf{V}(\mathbf{X}), \qquad (11)$$

130 with the initial condition of $\mathbf{X}(t_0;\mathbf{a}) = \mathbf{a}$, where ***V*** is Lagrangian velocity, which can be derived
131 from seepage velocity ***v***. Now we are interested on the travel time, the time taken for a particle to
132 travel from the initial position **a** to a well or to across a control plane that is perpendicular to the
133 mean flow direction and located at some distance from the source. The travel time $\tau = t - t_0$ can
134 be determined from (11) using the particle tracking technique.

135 In summary, in addition to direct measurements on soil properties, it is assumed that we also
136 have pressure head measurements at $m_\psi$ locations, water content measurements at $m_\theta$ locations,
137 and concentration measurements at $m_C$ locations (or travel time measurements at $m_t$ locations).
138 These indirect measurements can be taken at a number of different elapsed times. The aim is to
139 estimate statistics of parameter fields based on all these measurements.

140 **3. Representation of Soil Properties**

141 As mentioned above, the soil properties $p$, where $p = f$, $\beta$, or $\mu$, are treated as spatially stationary
142 random functions with mean $\langle p \rangle$, and covariance function $C_p(\boldsymbol{x},\boldsymbol{y})$. Because the number of
143 parameter values to be estimated is usually much larger than the number of available
144 measurements, it is often to parameterize the parameter fields. There are several different ways
145 to parameterize a parameter field (McLaughlin and Townley, 1996). In this study, these
146 parameter fields are represented by $n_p$ basis kernel functions $b_p(\mathrm{x},\chi)$ centered at some fixed
147 spatial locations $\chi_j^{(p)}, j = \overline{1,n_p}$

$$p(\mathbf{x}) = \sum_{j=1}^{n_p} \gamma_j^{(p)} b_p(\mathbf{x}, \boldsymbol{\chi}_j^{(p)}) \tag{12}$$

where $\gamma_j^{(p)}, j = \overline{1, n_p}$, are coefficients to be determined in the inverse procedure. The kernel functions can be chosen as, for example, an exponential function $b_p(\mathbf{x}, \boldsymbol{\chi}) = \exp\left(-\sum_{i=1}^{d} | x_i - \chi_i | / \lambda_i\right)$, where $d$ is the number of space dimensions and $\lambda_i$ is a parameter that controls the influence of the kernels $b_p(\mathbf{x}, \boldsymbol{\chi})$ in the $i^{\text{th}}$ dimension. For given basis kernel functions, the parameter field $p(\mathbf{x})$ can be computed from coefficients $\gamma_j^{(p)}, j = \overline{1, n_p}$. Note that the points $\chi_j^{(p)}, j = \overline{1, n_p}$, on which the kernel functions are based, may be chosen differently for different parameter fields.

## 4. Bayesian Inference

The essence of the Bayesian approach is Bayes' Theorem, which can be understood as a mathematical description of the learning process. Bayesian statistical inference requires an additional input not needed by frequentist procedures such as maximum likelihood: *a prior* probability distribution for the parameters $\gamma$, which embodies our judgment before seeing any data $D$ of how plausible it is that the parameters could have values in the various regions of the parameter space. The introduction of a prior is the crucial element that converts statistical inference into an application of probabilistic inference. When we combine a prior distribution $\pi(\gamma)$ with the conditional distribution for the observed data, we get a joint distribution:

$$\pi(\gamma, D) = \pi(\gamma)\pi(D \mid \gamma) = \pi(D)\pi(\gamma \mid D) \tag{13}$$

where $D = (f_o, \beta_o, \mu_o; \psi_o, \theta_o, \tau_o, C_o)$ includes all observed data: $f_o$, $\beta_o$, and $\mu_o$ are respectively the vectors of $f$, $\beta$, and $\mu$ measurements, and $\psi_o, \theta_o, \tau_o, C_o$ are vectors of measurements for pressure

167 head, water content, travel time, and solute concentration. From this we can derive Bayes' rule
168 for the posterior distribution of the parameters given observed data *D*:

169 $$\pi(\gamma \mid D) \propto L(D \mid \gamma)\pi(\gamma) \tag{14}$$

170 where $L(D|\gamma)$ is the likelihood function. For our problem described above, the likelihood function
171 $L(f_o,\beta_o, \mu_o; \psi_o,\theta_o, \tau_o ,C_o \mid \gamma)$ of observed data $D$ given parameters $\gamma$ may be written as

172 $$L(D \mid \gamma) \propto \exp\left\{-\frac{1}{2}\sum_{p=f,\beta,\mu}\left(p_o - p_m\right)^T \sum_p^{-1}\left(p_o - p_m\right) - \frac{1}{2}\sum_{h=\psi,\theta,\tau,C}\left(h_o - h_m\right)^T \sum_h^{-1}\left(h_o - h_m\right)\right\}$$

173 (15)

174 where $\sum_p^{-1}$ is an $m_p \times m_p$ matrix determined by observation errors and representativeness of
175 measurements for parameters $p = f$, $\beta$, and $\mu$, $\sum_h^{-1}$ is an $m_h \times m_h$ matrix accounting for
176 observation error and model discrepancy on dependent variables $h = \psi$, $\theta$, $\tau$, or $C$. Note that
177 some of terms in (15) can be missing, depending on the availability of observation data. For the
178 Bayesian approach, we need to specify a prior distribution for $\gamma$. One such example is

179 $$\pi(D \mid \gamma) \propto \lambda_\gamma^{m/2} \exp\left\{-\frac{1}{2}\lambda_\gamma \sum_{p=f,\beta,\mu}\sum_{i\sim j}\left(\gamma_i^{(p)} - \gamma_j^{(p)}\right)^2\right\} = \lambda_\gamma^{m/2} \exp\left\{-\frac{1}{2}\lambda_\gamma \left(\gamma^{(p)}\right)^T W \gamma^{(p)}\right\}$$

180 (16)

181 where $\sum_{i\sim j}$ is the set of pairwise adjacencies, and matrix $W$ is defined as

182 $$W_{ij} = \begin{cases} -1 & \text{if } i \text{ and } j \text{ are adjacent} \\ n_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

and $n_i$ is the number of neighbors to location $i$. The prior distribution for the hyperparameter $\lambda_\gamma$ in (16) can be chosen as a Gamma distribution

$$\pi(\lambda_\gamma) \propto \lambda_\gamma^{a-1} e^{-b\lambda_\gamma} \tag{18}$$

Finally, the posterior distribution of parameters ($\gamma$, $\lambda_\gamma$) given observed data $D$ can be written as

$$\pi(\gamma, \lambda_\gamma \mid D) \propto L(D \mid \gamma)\pi(\gamma \mid \lambda_\gamma)\pi(\lambda_\gamma)\,. \tag{19}$$

Estimation and inference are based on this posterior distribution. Note that we only need to know the posterior distribution up to a constant proportionality for our Markov Chain Monte Carlo simulations discussed in the next section.

## 5. Markov Chain Monte Carlo Simulations

Sampling methods based on Markov chains incorporate the required search aspect in a framework where it can be proved that the correct distribution is generated at least in the limit as the length of the chain grows. Writing $(\gamma, \lambda_\gamma)^{(t)}$ for the set of variables at time step $t$, where $\gamma$ itself is a vector, the chain is defined by giving an initial distribution $(\gamma, \lambda_\gamma)^{(0)}$ and the transition probabilities for $(\gamma, \lambda_\gamma)^{(t)}$ given the value for $(\gamma, \lambda_\gamma)^{(t-1)}$. These probabilities are chosen so that the distribution of $(\gamma, \lambda_\gamma)^{(t)}$ converges to that for $(\gamma, \lambda_\gamma)$ as $t$ increases and so that the Markov Chain can feasibly be simulated by sampling from the initial distribution and then in succession from the conditional transition distributions.

Typically the Markov chain explores the space in a "local fashion". In some methods for example $(\gamma, \lambda_\gamma)^{(t)}$ differs from $(\gamma, \lambda_\gamma)^{(t-1)}$ in only one component of the state, e.g., it may differ with respect to $\gamma_i^{(t)}$, a component of $\gamma$, for some $i$ but have $\gamma_j^{(t)} = \gamma_j^{(t-1)}$ for $j \neq i$. Other methods may

change all components at once but usually by only a small amount. Locality is often crucial to the feasibility of these methods. In the Markov chain framework it is possible to guarantee that such step-by-step local methods eventually produce a sample of points from the globally correct distribution. The procedure implemented in this study can be summarized as follows:

(1) Initialize parameters at some value $(\gamma, \lambda_\gamma)^{(0)}$. Theoretically, they can be initialized by any numbers taking from the initial distribution. For example, one can initialize $(\gamma, \lambda_\gamma)$ by drawing a set of random numbers. In this study, we choose $\gamma$ such that the initial parameter fields are close to the mean fields.

(2) Update each $\gamma_i$ according to Metropolis rules:

- Draw a value $\gamma_i^*$ from a uniform distribution $U[\gamma_i^{(t-1)} - r, \gamma_i^{(t-1)} + r]$, where $r$ is a pre-determined small number. Let $\gamma^*$ be a vector that differs from $\gamma^{(t-1)}$ only in their $i^{th}$ component, i.e., $\gamma^* = (\gamma_1^{(t-1)}, \ldots, \gamma_{i-1}^{(t-1)}, \gamma_i^{(t)}, \gamma_{i+1}^{(t-1)} \ldots, \gamma_n^{(t-1)})^T$.
- Compute $\eta = \pi(\gamma^*, \lambda_\gamma|D)/\pi(\gamma^{(t-1)}, \lambda_\gamma|D)$. Accept new value $\gamma_i^*$ with probability $\min(1,\eta)$, else reject new value $\gamma_i^*$ (i.e., keep $\gamma_i$ unchanged). In other words, if the newly proposed value increases the posterior probability (i.e., $\eta > 1$), the new value is accepted. Note that even if the proposed value reduces the posterior probability (i.e., $\eta < 1$), the value could still be accepted with a probability of $\eta$.

(3) Update $\lambda_\gamma$ given $\gamma$ according to the following posterior distribution of $\lambda_\gamma$, again using Metropolis rules:

$$\pi(\lambda_\gamma \mid \gamma) \propto \pi(\gamma \mid \lambda_\gamma)\pi(\lambda_\gamma) \sim \Gamma\left(a + \frac{m}{2}, b + \frac{\gamma^T W \gamma}{2}\right), \tag{20}$$

where $a$ and $b$ are two prescribed constants.

(4) Repeat steps 2 and 3 as needed.

There are considerable discussions on the convergence diagnosis of the MCMC algorithms (Brooks, 1998). For simplicity, the convergence of the chain in this study is determined from the plots of the log posterior density versus the number of updates. Because the first portion of the chain may depend on the initial setting, to reduce the possible effect of the starting values, the first portion of the chain is discarded (called burn-in or warm-up period). There are some discussions on the length of the burn-in period. In practice, throwing away the initial 1 or 2 % of runs will usually suffice (Geyer, 1992). Each realization in the remaining chain will fit the observed data very well and parameter statistics can be computed from this chain. Since the chain explores the parameter space in a local fashion, realizations in the chain are usually auto-correlated. To reduce the autocorrelation, subsampling technique is used in computing the parameter statistics from the chain (Geyer, 1992).

## 6. Illustrative Examples

In this section we first demonstrate our inverse method for one-dimensional unsaturated flow in a hypothetical heterogeneous porous medium, whose properties are assumed to be known and will be compared with the inversion results. The inverse method is then applied to real examples of the water-content data collected on the Pajarito Plateau, New Mexico.

### 6.1 Synthetic Example

For this hypothetical problem, the flow domain is a soil column with a depth of $L$ = 10 m, uniformly discretized into 100 elements (101 nodes). The pressure head is prescribed at the bottom as $\psi(0) = 0$ (the water table) and water infiltration with a rate of $q$ = 0.002 m/day is

prescribed at the top. The statistics of soil properties for this hypothetical soil are given as $\langle f \rangle$ = 1.0, $\sigma_f^2 = 1.0$, $\langle \beta \rangle = 0.5$, $\sigma_\beta^2 = 0.01$, $\langle \mu \rangle = -0.9$, $\sigma_\mu^2 = 0.01$, and a correlation length of $\lambda_f = \lambda_\beta = \lambda_\mu$ = 1.0. The variability of these parameters can also be given in terms of the coefficient of variation as $CV_{Ks}$ = 131%, $CV_\alpha = CV_n \approx$ 10%. We then generate three random fields as "true" parameter fields, using the specified statistics and exponential covariance functions for three soil properties. These "true" fields are used as references to access the quality of our inverse model. We solve flow equations (both steady state and transient flow) and transport equations using these true parameter fields to obtain "true" head fields (steady state or transient) and concentration fields.

We take $n_f = n_\beta = n_\mu = 5$ samples from these true parameter fields as our direct measurements of $f$, $\beta$, and $\mu$. We also take $n_\psi = n_\theta = 20$ samples for the pressure head and water content and $n_C$ = 6 concentration samples at three elapsed times $t$ = 0.1, 10.0, and 100. The measurement locations are illustrated in Figure 1. After taking all these measurements, we proceed as though the ensemble statistics (the mean, variance, and correlation lengths) used in generating these original parameter fields are not available, and that all we have are the direct and indirect measurements. Our purpose is to estimate three parameter fields using these measurements.

As a first step, we may need to estimate sample statistics of soil properties. Several methods can be used to estimate the sample statistics, i.e., the mean, variance, and correlation length. One simplest way is to compute the mean and the variance from direct measurements and find the correlation length by fitting the variogram. An alternative is to estimate these statistics from the maximum likelihood method using both direct and indirect measurements. In our Markov Chain Monte Carlo method (MCMC), these statistics can be estimated simultaneously in the inverse

process. However, in this preliminary study we compute these statistics from direct measurements only. These estimates are $\langle f \rangle = 1.711$, $\sigma_f^2 = 1.455$, $\langle \beta \rangle = 0.477$, $\sigma_\beta^2 = 0.012$, $\langle \mu \rangle = -0.848$, $\sigma_\mu^2 = 0.008$, and a correlation length of $\lambda_f = \lambda_\beta = \lambda_\mu \approx 1.2$.

For the MCMC method, based on the domain size and the estimated correlation length of about 1.2, we use a grid of 18 basic kernel locations, more-or-less uniformly distributed in the domain as shown in Figure 1. The error matrices $\sum_p$ are chosen to be $\varepsilon_p I_{n_p}$, where $p = f, \beta, \mu, \psi, \theta$, or C, and $I_n$ stands for an identical matrix of $n \times n$, and ε's are prescribed standard deviations for errors of variables $p$'s. Here we choose $\varepsilon_f = 0.10$, $\varepsilon_\beta = 0.02$, $\varepsilon_\mu = 0.02$, $\varepsilon_\psi = 0.01$, $\varepsilon_\theta = 0.005$, and $\varepsilon_C = 0.002$.

The estimated soil parameter fields from the MCMC method (dashed curves) are illustrated in Figure 2, as compared to the true parameters fields (solid curves). It is seen from the figure that the estimated parameter fields match the trend of the true fields very well. Note that the estimated values at the conditioning points deviate from their corresponding true values, because the specified measurement errors in the MCMC method allow the estimated values vary within some ranges. The degree of such deviations is characterized by the standard deviation of errors specified by $\varepsilon_p$, where $p = f, \beta$, or $\mu$.

Figures 3, 4, and 5 compare the true pressure head, moisture content and concentration profiles against the simulated ones at three different times. Although the estimated fields cannot capture the detail variation of true fields, they reproduce the general trends of the true fields very well. These results suggest that the MCMC method is capable of providing excellent fits to the

hydrologic data, making it an appropriate method for the water-content data for the Pajarito Plateau.

### 6.2 Application of the MCMC method to the Los Alamos site, New Mexico

In this section, we apply the MCMC method described above to use the water-content data to estimate local infiltration rate at a number of well locations in the vicinity of the LANL site. Since there is significant model development required to implement this new analysis approach, we begin by discussing the model setup. Then, we examine one of the wells, MCOBT-4.4, in greater detail, using it to assess the effectiveness of the technique. We then present the analyses for the other wells in a more concise way.

Table 1 lists the hydrologic properties used for the model (Rogers and Gallaher, 1995; Rogers et al., 1996; Broxton et al., 2002). The table contains both the permeability and porosity values used for each unit, as well as unsaturated hydraulic parameters, $\alpha$ and n, for defining the van Genuchten (1980) constitutive relationship. We assume that porosity is a deterministic constant (a constant for each type of stratigraphic unit), while the permeability, pore-size distribution parameter $\alpha$, and fitting parameter n, are spatially random functions and are modeled by log-normal distributions. The property values of a unit listed in the table will be used as initial values for the MCMC simulation if the unit occurs in a borehole.

#### 6.2.1 Model Setup

Although the method is not restricted to these assumptions, the simplified model used to perform the MCMC analyses of the moisture content profiles is a one-dimensional, steady state flow model with uniform numerical grids. The goal of the analysis is to estimate the infiltration rate and hydrologic parameters associated with the measured water contents from the wells.

Therefore, in general, the infiltration rate, the saturated hydraulic conductivity, and the van Genuchten parameters are jointly varied in the inverse model runs. For testing purposes, other strategies are employed to examine the influence of these choices on the inversion results.

The finite-element heat and mass-transfer code (FEHM) of Zyvoloski et al (1997) is used to perform the model runs. Grid spacing was chosen so that, in general, the spacing was smaller than the spacing of water content measurements. This approach reduces the possible loss of conditional points. For cases in which more than one measurement is located in single computational grid cell, the average water content value of all measurement points were taken this node as a new conditional point. In addition, since for each node there are three parameters to be estimated, there are practical limitations to the resolution of the grid for situations in which parameters are allowed to vary on a node-by-node basis. For these reasons, we selected a grid spacing ranging from about 3 to 5 feet, depending on the problem.

The one-dimensional column representing a given well is divided into a number of zones, mainly based on the hydrostratigraphy determined from well logs. When detailed information (upper and lower bounds) about subdivisions in a particular formation is available, each subdivision is defined as an individual zone. Initially, hydraulic properties are assumed to be the same for all subdivisions of the formation, and the MCMC method seeks to fit the water content data by adjusting the hydrologic parameters in the column. Three strategies are employed for this purpose:

1. Properties are assumed to be random constants within a zone, but vary from zone to zone.
2. Properties are defined by a set of kernel functions, the coefficients of which are updated sequentially.

3. Properties are defined on a node-by-node basis.

These methods represent increasingly complex models of the heterogeneities in hydrologic properties. Part of our study will consist of assessing which method is the appropriate level of complexity for a given data set, taking into consideration data sparseness and computational efficiency. In most cases, we will not use the node-by-node approach (method 3) due to computational costs and data limitations. For example, for a soil column of 100 nodes, there are 300 hydraulic parameters plus the infiltration rate, which means that for any particular parameter, 301 model runs are required to perform an update. The reason to develop method two is that it seems like a reasonable compromise that allows heterogeneities within thick units to be modeled without an excessive number of parameters.

Finally, because the water content measurements are usually available in the upper part of the well and hydraulic properties in the deep zones that are far away from the measurement locations have little impact on data fitting, properties in these deep zones are modeled as random constants.

In each MCMC simulation, the run starts from an initial set of soil parameters and infiltration rate, and the parameter values are updated sequentially based on the rules described above. One of the virtues of the approach is that the parameter uncertainty statistics can be derived from the variability of values obtained during the chain. However, the estimated parameters in the first part of the chain strongly depend on the choice of initial settings. Thus, this initialization phase of the simulation (called the burn-in period) is ignored when computing parameter statistics. The length of burn-in, which is problem dependent, is determined graphically from a plot of the negative log posterior versus the number of updates. Initially, this metric is possibly very large,

since the initial setting may significantly deviate from the true solution. The burn-in period is approximated graphically as the point in the simulation at which the negative log posterior becomes stabilized.

Since the MCMC method takes a perturbation approach to parameter updating (i.e., the new value for any parameter is derived by adding a possible perturbation to the old value), the sequence of values for each parameter are not completely independent. As a consequence, the computed parameter variance may be artificially small. Because we seek to use the analysis to compute the statistics (mean and variance) for each parameter, we take a subset of the sequence by selecting values within a predefined interval. The variance for parameter is then used to construct the confidence intervals around the mean predictions.

### 6.2.2 Borehole MCOBT4-4

Geologic units encountered in MCOBT-4.4 consist of the following, in descending order: canyon-bottom alluvium; deposits of the Cerro Toledo interval; the Otowi Member of the Bandelier Tuff, including the basal Guaje Pumice Bed; an upper sequence of fanglomerate and sand deposits of the Puye Formation; lavas, interflow units, and subflow deposits of the Cerros del Rio volcanic field; and a lower sequence of fanglomerate deposits of the Puye Formation. Canyon-bottom alluvium (Qal) was cored from 0 to 63.7 ft depth at MCOBT-4.4. The alluvium consists predominantly of moderately weathered detritus of the Tshirege Member of the Bandelier Tuff and is unconsolidated. In this study, we exclude this layer of alluvium in our simulations. The simulation domain ranges from the depth of 64ft (elevation 6769.2ft, or 2063.25m) to 493ft (elevation 6343.2ft, or 1933.4m), where perched water occurs. The soil column is uniformly discretized into 143 elements of size 3ft (0.914m). There are 30

375 measurements of water content available. Here we assume that the error matrix for the water
376 content as appeared in (15) is a diagonal $\sum_{\theta} = \varepsilon_{\theta}^2 I_{n_\theta}$ with $\varepsilon_{\theta} = 0.01$.

377 We conducted several simulation runs to investigate the sensitivity of simulation results on
378 model settings. In the first case, we fix the hydrologic properties of the column based on
379 stratigraphic units at the borehole, while allowing the infiltration rate to vary, starting from a
380 lower value of $5\times10^{-7}$ kg/sec (=15.7 mm/year) with a possible maximum increment of $r = 2\times10^{-7}$
381 kg/sec at each update. The fitting between the observed and modeled water content is not very
382 good for this run, as evidenced by a large root-mean-square-error (RMSE) of water content
383 (0.056) in Table 2. With only the infiltration rate being varied, and no adjustment of the
384 hydrologic properties, the method is limit in its ability to capture the details of the water content
385 profile.

386
387 Next, we examine the results with an identical parameter strategy as above, but start the MCMC
388 simulation with a very high infiltration rate of $5\times10^{-5}$ kg/sec (=1568 mm/year) and a possible
389 maximum increment of $r = 1\times10^{-6}$ kg/sec at each update. Note that the estimated infiltration rate
390 from these runs are very close, indicating that final infiltration rate is independent of the initial
391 value. However, it seems from these two cases that good fits to the measured water content data
392 cannot be obtained by simply varying the infiltration rate alone. Modification of hydraulic
393 parameters appears to be required to fit the observed water-content data.

394 In the third case, in addition to the variable infiltration rate, we allow the hydraulic properties to
395 vary as random constants, i.e., all three soil parameters for each layer varying in probability
396 space but being uniform in the layer. Again, it seems that the fitting is not that good (Table 2).
397 All these three cases indicate that some of the layers at well MCOBT4.4 have to be modeled by

individual nodes or by sets of kernel functions such that the soil properties vary in the layers rather than uniform layers as in the previous three cases.

In the next run (case 4), we model the soil properties in Qbo (depth 101.9 to 462 ft) by 10 kernel functions while other three thin layers (Oct, 64-101.9ft, Qbog, 462-474ft, and Tpf, 474-493 ft) by individual nodes. We have in total 143 grid nodes, 23 of which are modeled individually and a 121 nodes of which are represented by 10 kernel functions. This ends up 33 groups of soil parameters, plus one additional parameter for the infiltration rate, which means that each parameter will be updated after 100 (= 3 × 33 + 1) model runs. The initial soil properties at grid nodes are assigned based on their stratigraphic units and the initial infiltration rate is $5\times10^{-7}$ kg/sec (=15.7 mm/year). Figure 6(a) depicts the negative log posterior as a function of the number of updates. The figure shows that the negative log posterior reduces quickly at the beginning as the number of updates increases and then stabilizes at about 10, which is equivalent to the root-mean-square-error (RMSE) of 0.0082 for water content. Figure 6(b) illustrates the decrease of the negative log posterior for the first hundreds of updates as a function of 34 parameter groups, where the zeroth variable corresponds to the infiltration rate and the first group of variables represents soil properties at the top node, and so on. The figure clearly shows that updates on the infiltration rate have a significant effect on reducing the negative log posterior. Furthermore, varying the soil properties in the bottom part of the column does not have significant impacts on data fitting.

The comparison between the modeled and observed water content is illustrated in Figure 7(a), where results from different number of updates are also shown. The figure indicates that, by updating soil properties at each node the upper part of column, we are able to fit observed water

content very well. It is also seen from the figure that in the lower part of Qbo the water content could vary significantly, partially because there is no data in this part of the column. It is interesting to notice that the water content in the bottom two layers (below 1945m) changes very little, even though the soil properties in these layers are highly heterogeneous (Fig. 8). The estimated mean and standard deviation of the infiltration rate are 125.3 mm/yr and 30.9 mm/yr, respectively. Since the infiltration rate is normally distributed (as shown later), these statistics mean that at the 95% confidence level the actual infiltration rate ranges from 63.5 mm/yr to 187.1 mm/yr.

All simulations presented thus far have used the mean values of parameters listed in Table 1 for the initial values of the MCMC run. To test the sensitivity of the inversion results on initial hydrologic properties, three more cases were performed with different starting parameter values than those in case 4, which is considered to be the base case. In case 5, we initialize the model with permeability values 20 times larger than the values based on the stratigraphic units. In cases 6 and 7, the initial $\alpha$ and n fields are respectively 1.5 larger than those in case 4.

The estimated mean and standard deviation of infiltration rate for these different cases (cases 4-7) are listed in Table 2, along with the root-mean-square-error (RMSE) for each case. Although the statistics of the infiltration rate differ from case to case, all four cases (the base case and the three sensitivity runs) fit the observed water content equally well, as evidenced from the RMSE in Table 2. The histograms of the infiltration rate for these cases are illustrated in Figure 9. For comparison purposes, the histogram for the base case is also shown. It is interesting to see that the histogram of the infiltration rate for the base case and the case with an initially high permeability setting follows approximately the normal distribution, whereas the cases with high

α and n values the histogram exhibit skewness and, in the case of Figure 9b, bimodality. In addition, the table shows that the infiltration statistics from the case with an initially high permeability setting (case 7) are very close to those of the base case, indicating that initial setting for the permeability field does not have a significant impact on the data fitting. On the other hand, initially relatively large deviations of α and *n* from their mean values yield non-Gaussian distributions of the infiltration rate. This may stem from the fact that the infiltration is linearly proportional to permeability, while its relationships with α and *n* are nonlinear.

### 6.2.3 Other Boreholes

The MCMC method has been applied to several other wells at the Los Alamos site to estimate the infiltration rate at the well locations. The summary of infiltration rates presented in Table 3 illustrates the wide range of infiltration values obtained from water content profiles, depending on the topographic setting and location within canyons. Overall, the results are in general agreement with past analyses of infiltration rates at these and other similar locations on the Pajarito Plateau. For example, the difference in infiltration rate estimated between LADP-3 and LADP-4 illustrates the vast difference in downward percolation flux depending on whether the location in a wet canyon or a mesa/dry canyon. Within Los Alamos Canyon itself, the difference between LADP-3 and LAOI(A)-1.1 is thought to be due to the proximity of the latter well to a more intensely fractured region associated with the Guaje Mountain fault zone (Gray, 1997). The estimated value in R-9, located further down canyon, is lower than that of either of these two wells. However, the infiltration into highly fractured basalts with low permeability matrix may violate the basic assumption of porous flow in some of the rock units, making this estimate potentially suspect.

In Mortandad Canyon, the relative infiltration rates in the two wells are in keeping with the locations of these wells. MCOBT-4.4 is located further up canyon, and it may be that the source of water in MCOBT-8.5 may be depleted by infiltration and ET processes at this down-canyon location. Another factor to consider in Mortandad Canyon is that the analyses in these intermediate wells are focused on water content values in the upper part of the stratigraphic section. Recent reductions in the water discharged from the Radioactive Liquid Waste Treatment Facility (RLWTF) may have already impacted the water content values in the upper parts of the vadose zone where the measurements are made. If this is the case, then the infiltration rate estimates reflect the present-day infiltration rate, rather than the historical, presumably higher, infiltration rate. This concept of declining infiltration rate over time due to changes in the operation of the RLWTF has been promoted by Kwicklis et al. (2005), who applied the concept to an interpretation of several tritium peaks in well R-15. In principle, a transient analysis of the water content information could be performed for these wells to examine this possibility, but this approach was beyond the scope of this study. The analysis developed herein should be extended to this and other wells in Mortandad Canyon to evaluate the water content and contaminant profile information.

## 7. Summary and Conclusions

With regard to the new analysis technique, the MCMC method proved to be a very effective means for determining the mean and standard deviation of infiltration from the water content profiles. Obtaining the standard deviation is an important advance because many inverse techniques yield unrealistic estimates of the uncertainty. The MCMC method should provide appropriate uncertainty estimates because it takes into account heterogeneity in the medium properties, using that as a basis for obtaining better fits to the water content data. Numerical

experiments with the method illustrate that including small-scale heterogeneity is important for improved matches to the data. When applied to the data from wells in the vicinity of LANL, the method provided reasonable estimates of uncertainty for wells in a variety of topographic settings. The more advanced versions of the method that include transient flow and solute transport should be useful for interpreting data currently being collected as part of the ER activities in Mortandad Canyon and other locations around the Laboratory.

## Acknowledgments

This work was conducted under the auspices of the U.S. Department of Energy, Los Alamos Groundwater Protection Program (GWPP).

## References

Birdsell, K. H., B. D. Newman, D. E. Broxton, and B. A. Robinson, 2005. Conceptual models of vadose-zone flow and transport beneath the Pajarito Plateau, Los Alamos, New Mexico. *Vadose Zone J.*., 4: 620-636.

Brooks S. P, 1998. Markov chain Monte Carlo method and its applications, Statistician, 47, 69-100.

Brooks, R. H., and A. T. Corey, 1964. Hydraulic properties of porous media, *Hydrol. Pap.* 3, Colo. State Univ., Fort Collins.

Broxton, D., D. Vaniman, P. Longmire, B. Newman, W. Stone, A. Crowder, P. Schuh, R. Lawrence, E. Tow, M. Everett, R. Warren, N. Clayton, D. Counce, E. Kluk, and D. Bergfeld, 2002. Characterization Well MCOBT-4.4 and Borehole MCOBT8.5 Completion Report, Los Alamos National Laboratory report LA-13993-MS, Los Alamos, New Mexico.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten, 2005. Inverse problem in hydrogeology, Hydrogeol J., 13, 206-222.

Chen, J. S. Hubbard, Y. Rubin, C. Murray, E. Roden, and E. Majer, 2004. Geochemical characterization using geophysical data and Markov Chain Monte Carlo methods: A case study at the South Oyster bacterial transport site in Virginia, *Water Resour. Res*., 40, W12412, doi:10.1029/2003WR002883.

517 Dander, D. C., 1998. Unsaturated groundwater flow beneath upper Mortandad Canyon, Los
518 Alamos, New Mexico, Los Alamos National Laboratory report LA-UR-98-4759, Los
519 Alamos, New Mexico.

520 Efendiev, Y., A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallick, 2005. An efficient two-stage
521 Markov chain Monte Carlo method for dynamic data integration, Water Resour. Res., 41,
522 W12423, doi:10.1029/2004WR003764.

523 Gardner, W.R., 1958. Some steady state solutions of unsaturated moisture flow equations with
524 application to evaporation from a water table, *Soil Sci.,* 85, 228-232.

525 Geyer, C. J., 1992. Practical Markov Chain Monte Carlo, Statistical Science, 7(4), 473-483.

526 Ginn, T. R., J. H. Cushman, 1990. Inverse method for subsurfacr flow: a critical review of
527 stochastic techniques, *Stoch. Hydrol. Hydraul*., 4, 1-26.

528 Gray, R. N., 1997. Hydrologic budget analysis and numerical simulations of groundwater flow in
529 Los Alamos canyon near Los Alamos, New Mexico, Master's Thesis, University of New
530 Mexico, Albuquerque, New Mexico

531 Kwicklis, E., M. Witkowski, K. Birdsell, B. Newman, and D. Walther, 2005. Development of an
532 infiltration map for the Los Alamos area, New Mexico, Vadose Zone J., 4, 672-693.

533 Li, B., and T. J. Yeh, 1999. Cokriging estimation of the conductivity field under variably
534 saturated flow conditions, Water Resour. Res., 35(12), 3663-3674.

535 Lu, Z., and D. Zhang, 2002. Stochastic analysis of transient flow in heterogeneous variably
536 saturated porous media: the van Genuchten-Mualem constitutive model, *Vadose Zone*
537 *Journal*, 1, 137-149.

538 Lu, Z., D. Higdon, and D. Zhang, 2004. A Markov Chain Monte Carlo method for groundwater
539 inverse problem, *Developments in Water Resources*, 55, 1273-1283.

540 McLaughlin, D., and L. R. Townley, 1996. A reassessment of the groundwater inverse problem,
541 *Water Resour. Res*., 32(5), 1131-1162.

542 Oliver, D. S., L. B. Cunha, and A. C. Reynolds, 1997. Markov chain Monte Carlo methods for
543 conditioning permeability field to pressure data, Math. Geol., 29(1), 61-91.

544 Rogers, D. B., and B. M.Gallaher, 1995. The unsaturated hydraulic characteristics of the
545 Bandelier Tuff, Los Alamos National Laboratory report LA-12968-MS, Los Alamos,
546 New Mexico.

547 Rogers, D.B., B.M. Gallaher and E.L. Vold, 1996. Vadose zone infiltration beneath the Pajarito
548 Plateau at Los Alamos National Laboratory," New Mexico Geological Society
549 Guidebook, 47th Field Conference, Jemez Mountains Region, pp. 413-420.

550 Russo, D., 1988. Determining soil hydraulic properties by parameter estimation: On the selection
551 of a model for the hydraulic properties, *Water Resour. Res.,* 24, 453-459.

552 Russo, D., and M. Bouton, 1992. Statistical analysis of spatial variability in unsaturated flow
553 parameters, *Water Resour. Res.,* 28(7), 1991-1925.

554 van Genuchten, M. T., 1980. A closed-form equation for predicting the hydraulic conductivity of
555 unsaturated soils, *Soil Sci. Soc. Amer. J.*, 44 892- 898.

556 Vrugt, J. A., G. Schoups, J. W. Hopmans, C. Young, W. W. Wallender, T. Harter, and W.
557 Bouten, 2004. Inverse modeling of large-scale spatially distributed vadose zone
558 properties using global optimization, *Water Resour. Res*., 40, W06503, doi:10.1029
559 /2003WR002706.

560 Yeh, T. J., and J. Simunek, 2002. Stochastic fusion of information for characterizing and
561 monitoring the vadose zone, *Vadose Zone Journal*, 1, 207-221.

562 Yeh, T. J., and J. Zhang, 1996. A geostatistical inverse method for variably saturated flow in the
563 vadose zone, *Water Resour. Res*., 32(9), 2757-2766.

564 Yeh, T. J., S. Liu, R. J. Glass, K. Baker, J. R. Brainard, D. Alumbaugh, and D. LaBrecque, 2002.
565 A geostatistically based inverse model for electrical resistivity surveys and its
566 applications to vadose zone hydrology, *Water Resour. Res*., 38(12), 1278,
567 doi:10.1029/2001 WR001204.

568 Yeh, W. W., G., 1986. Review pf parameter identification procedures in groundwater hydrology:
569 The inverse problem, *Water Resour. Res*., 22(2), 95-108.

570 Zimmerman, D.A., G. de Marsily, C. A. Gotway, M. G. Marietta, C. L. Axness, R. Beauheim, R.
571 Bras, J. Carrera, G. Dagan, P. B. Davies, D. P. Gallegos, A. Galli, J. Gomez-Hernandez,
572 S. M. Gorelick, P. Grindrod, A. L. Gutjahr., P. K. Kitanidis, A. M. Lavenue, D.
573 McLaughlin, S. P. Neuman, B. S. Ramarao, C. Ravenne, and Y. Rubin, 1998. A
574 comparison of seven geostatisically based inverse approaches to estimate transmissivities
575 for modeling advective transport by groundwater flow, *Water Resour. Res.* 34(6),
576 p.1373.1413.

577 Zyvoloski, G.A., B.A. Robinson, Z.V. Dash, and L.L. Trease, 1997. Summary of the models and
578 methods for the FEHM application - A finite-element heat- and mass-transfer code.
579 Report LA-13307-MS, Los Alamos National Laboratory, Los Alamos, New Mexico.

580

581 Table 1. Stratigraphic Units and Their Hydraulic Properties at the Los Alamos Area

582

| No | Description | k($m^2$) | $\alpha$ (1/m) | n | porosity |
|---|---|---|---|---|---|
| 1 | Tsfu (all units below Tsfuv, undifferentiated, volcaniclastic) | 2.65E-13 | 5 | 2.68 | 0.35 |
| 2 | Tb1A (Cerros del Rio basalt-occurs within Santa Fe group) | 2.96E-13 | 5 | 1.5 | 0 |
| 3 | Tsfuv (Santa Fe Group, aquifer unit) | 2.65E-13 | 5 | 2.68 | 0.35 |
| 4 | Tb2 (Cerros del Rio basalt-occurs within Tsfuv) | 2.96E-13 | 5 | 1.5 | 0 |
| 5 | Tpt (Puye Formation, Totavi equivalent) | 4.73E-12 | 5 | 2.68 | 0.35 |
| 6 | Tpf (Puye Formation, fanglomerate) | 4.73E-12 | 5 | 2.68 | 0.35 |
| 7 | Tb4 (Cerros del Rio basalt-occurs within Puye Formation) | 2.96E-13 | 5 | 1.5 | 0 |
| 8 | Tt1 (Tschicoma dacite-occurs within Puye Formation) | 2.96E-13 | 5 | 1.5 | 0 |
| 9 | Tt2 (Tschicoma dacite-occurs within Puye Formation) | 2.96E-13 | 5 | 1.5 | 0 |
| 10 | Qbog (The Guaje pumice bed) | 1.53E-13 | 0.081 | 4.026 | 0.667 |
| 11 | Qbof (Otowi member of Bandelier Tuff) | 7.25E-13 | 0.66 | 1.711 | 0.469 |
| 12 | Qct (Cerro Toledo interval) | 8.82E-13 | 1.52 | 1.506 | 0.473 |
| 13 | Qbtt (Basal Pumice Unit, Tshirege member of Bandelier Tuff) | 1.01E-12 | 1.52 | 1.506 | 0.473 |
| 14 | Qbt1g(Glassy unit, Tshirege member of Bandelier Tuff) | 3.68E-13 | 2.22 | 1.592 | 0.509 |
| 15 | Qbt1v(Vitric unit, Tshirege member of Bandelier Tuff) | 1.96E-13 | 0.44 | 1.66 | 0.528 |
| 16 | Qbt2 (Unit 2, Tshirege member of Bandelier Tuff) | 7.48E-13 | 0.66 | 2.09 | 0.479 |
| 17 | Qbt3 (Unit 3, Tshirege member of Bandelier Tuff) | 1.01E-13 | 0.29 | 1.884 | 0.469 |
| 18 | Qbt3t (Unit 3t, Tshirege member of Bandelier Tuff) | 5.10E-13 | 2.57 | 1.332 | 0.466 |
| 19 | Qbt4 (Unit 4, Tshirege member of Bandelier Tuff) | 9.18E-14 | 0.667 | 1.685 | 0.478 |
| 20 | Qbt5 (Unit 5, Tshirege member of Bandelier Tuff) | 1.43E-14 | 0.17 | 1.602 | 0.349 |

583

584

585 Table 2. Test Cases for Well MCOBT-4.4
586

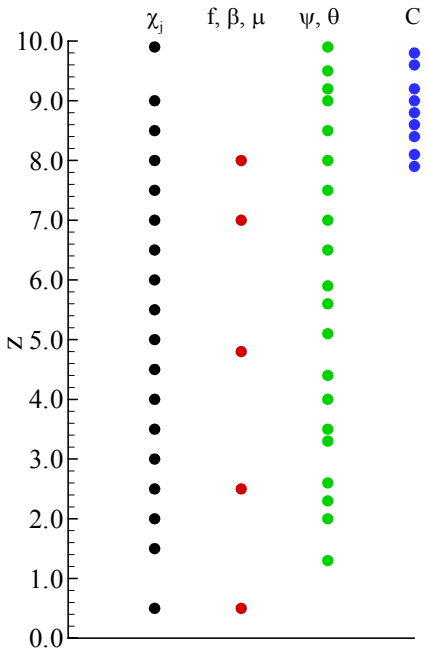
| Case # | Description | Mean infiltration rate (mm/yr) | Std deviation of infiltration rate(mm/yr) | RMSE of water content | Comments |
|---|---|---|---|---|---|
| 1 | Deterministic hydrologic properties; random infiltration rate (starting from lower q) | 442.6 | 25.6 | 0.0562 | Poor fitting |
| 2 | Deterministic hydrologic properties; random infiltration rate (starting from higher q) | 439.4 | 22.4 | 0.0562 | Poor fitting |
| 3 | Random constant hydrologic properties; random infiltration rate | 122.9 | 78.7 | 0.0488 | Normally distributed infiltration rate |
| 4 | Correlated hydrologic properties; random infiltration rate, starting from mean properties (base case) | 125.3 | 30.9 | 0.0084 | Normally distributed infiltration rate |
| 5 | Correlated hydrologic properties; random infiltration rate, starting from higher $\alpha$ values (1.5 times) | 44.3 | 11.8 | 0.0089 | |
| 6 | Correlated hydrologic properties; random infiltration rate, starting from higher n values (1.5 times) | 78.4 | 21.3 | 0.0098 | |
| 7 | Correlated hydrologic properties; random infiltration rate; starting from higher Ks values (20 times) | 129.2 | 28.2 | 0.0090 | Normally distributed infiltration rate |

587
588
589 Table 3. Summary of Estimated Infiltration Statistics at Different Wells
590

| Wells | Locations | Mean infiltration rate (mm/yr) | Std deviation of infiltration rate(mm/yr) | Comments |
|---|---|---|---|---|
| MCOBT-4.4 | MC Canyon | 125.3 | 30.9 | |
| MCOBT-8.5 | MC Canyon | 21.9 | 5.7 | No perched water |
| LADP-3 | LA Canyon | 220.0 | 16.7 | |
| LADP-4 | DP Canyon | 10.9 | 1.9 | No perched water |
| R-9 | LA Canyon | 139.9 | 84.2 | |
| LAOI(A)-1.1 | LA Canyon | 523.7 | 45.9 | |

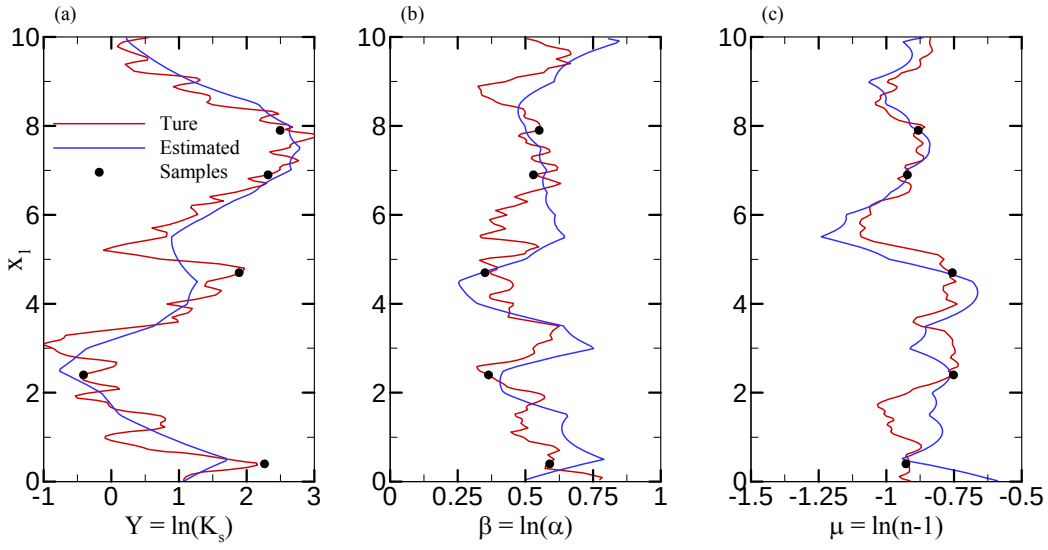
591
592
593
594

595 **List of Figures**




596

597 Figure 1. Layout of the problem configuration for the hypothetical example.




598

599 Figure 2. Comparison of the true soil properties and inverse results for the
600 synthetic example.
601

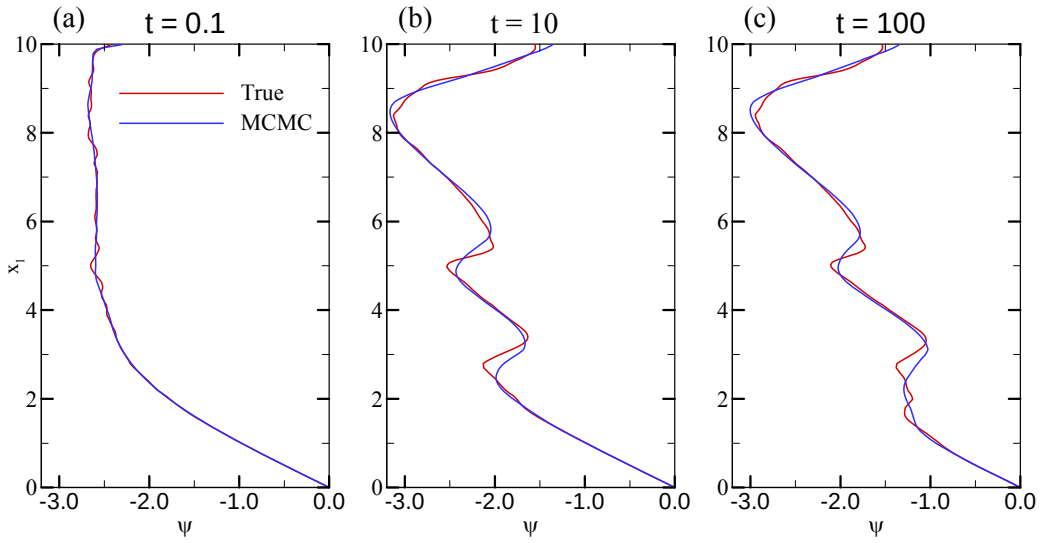



Figure 3. Comparison of the true and simulated pressure head profiles for the synthetic example.

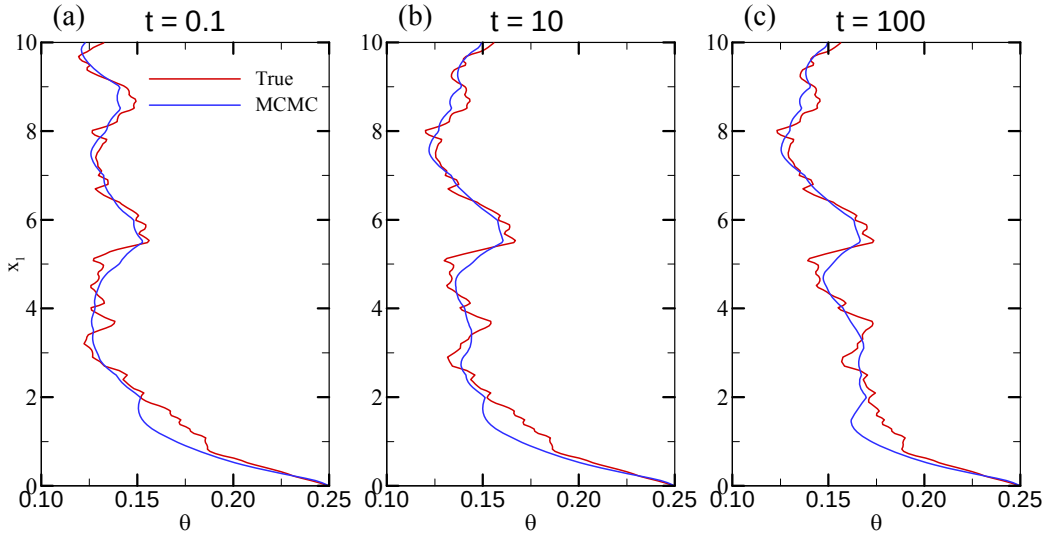



Figure 4. Comparison of the true and simulated moisture content profiles for the synthetic example.

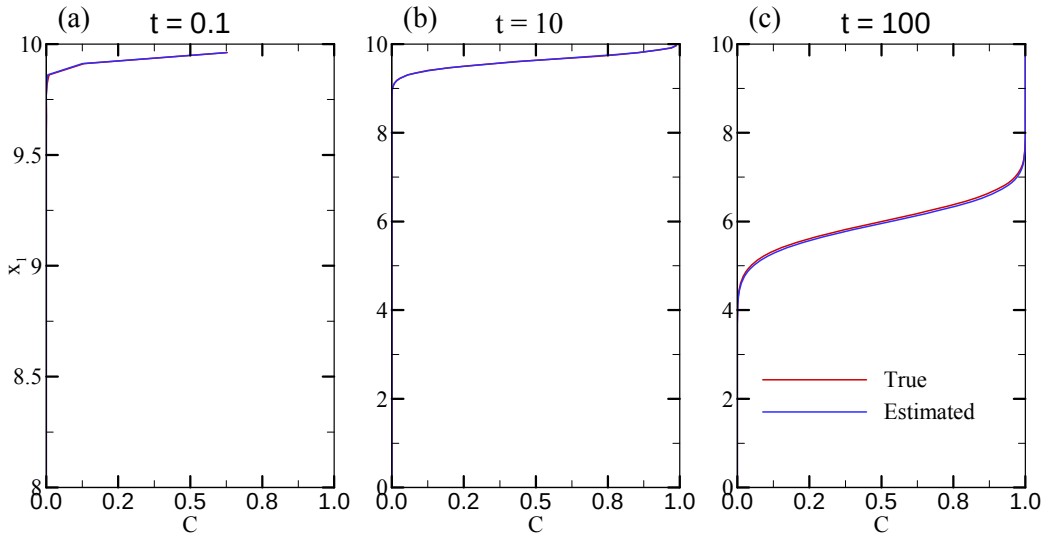



Figure 5. Comparison of the true and simulated concentration profiles for the synthetic example.

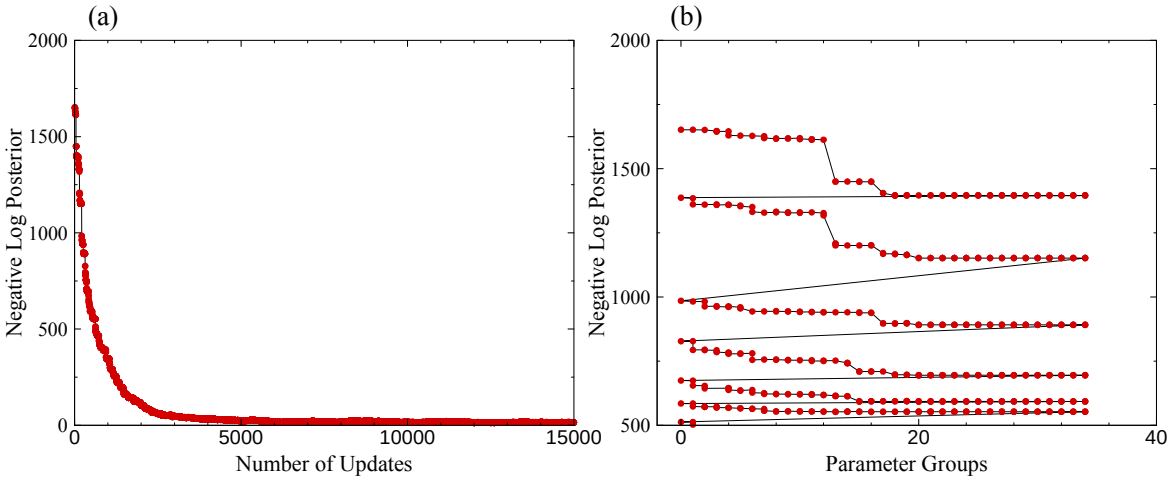



Figure 6. Negative log posterior vs (a) the number of updates, and (b) parameter groups.

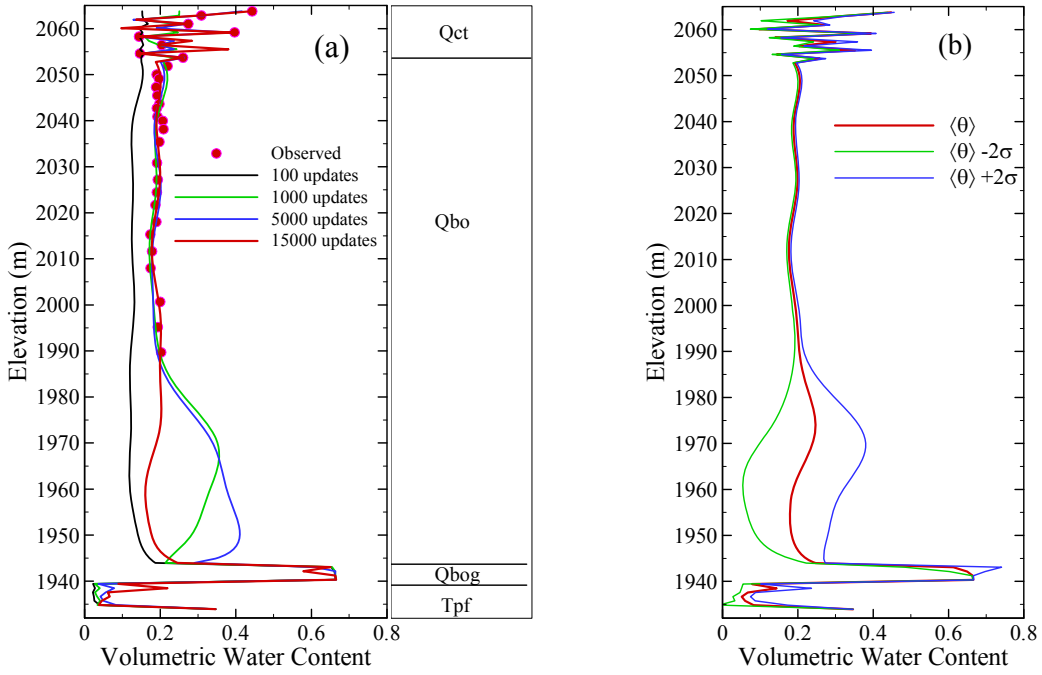



Figure 7. (a) Comparison of observed water content with fitted profiles of different numbers of updates for the case 4, and (b) the mean and confidence interval for volumetric water content.

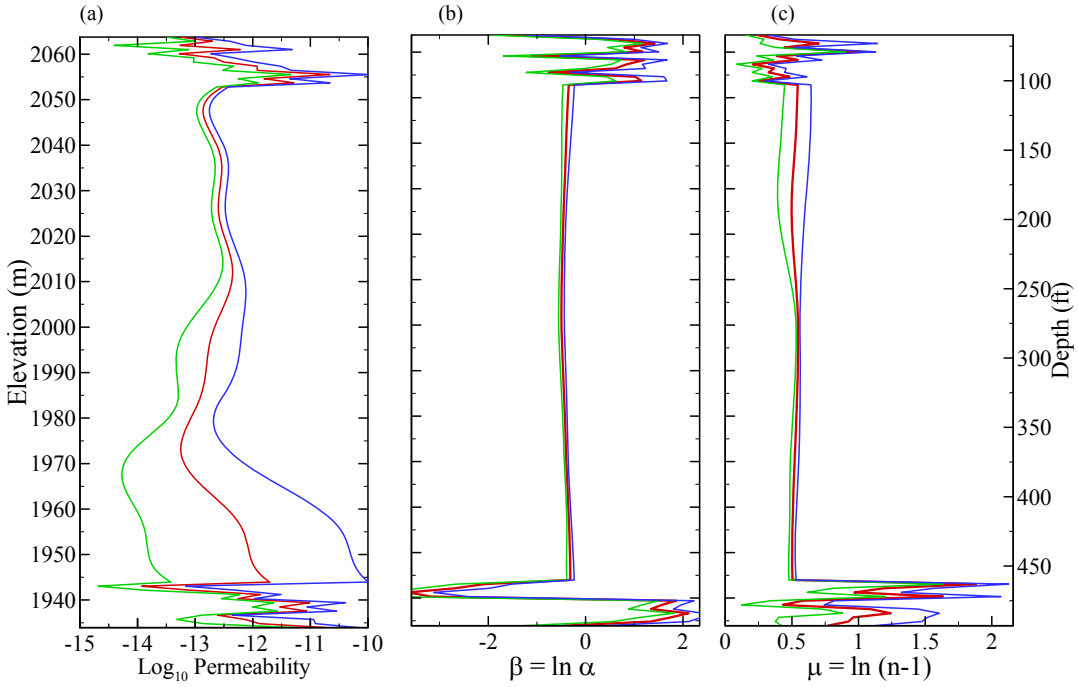



Figure 8. Estimated statistics of hydrologic properties, mean (red curves) and plus/minus two standard deviations (blue and green curves), for the case with random constant soil properties and a random infiltration rate.

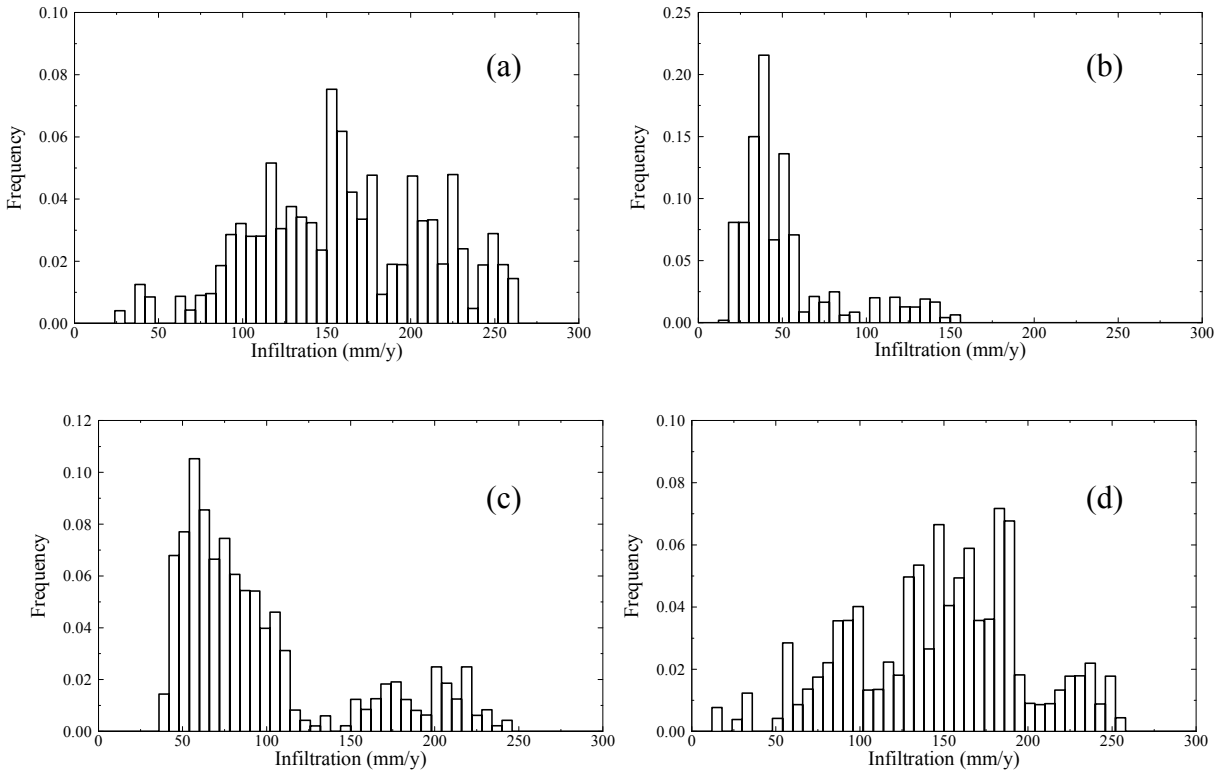



Figure 9. Histograms of the infiltration rate derived from simulations with various initial settings that are different from those based on rock units: (a) based on rock units, case 4, the base case, (b) α values are 1.5 times larger, case 5; (c) n values are 1.5 times larger, case 6; and (d) permeability is 20 times larger, case 7.